ADP Synthesis

Ralph G. Keyser
Senior Member of Technical Staff
Sandia National Laboratories
Albuquerque, New Mexico, USA

Larry S. Walker
Manager, Seismic Verification
Sandia National Laboratories
Albuquerque, New Mexico, USA

◆ Introduction

At the heart of the monitoring system for the Comprehensive Test Ban Treaty (CTBT) will be an Automated Data Processing (ADP) system charged with sorting through vast quantities of data from a world-wide network of sensors and providing distilled sets of information for decision makers. This system will evolve from the monitoring system prototypes in place today, but significant amounts of work remain to be done in order to complete this evolution.

This paper will address some of the challenges in the ADP area and the research efforts addressing those challenges. Efforts are underway at a number of agencies and organizations, aimed at successfully meeting the challenges presented by an automated data processing system for a CTBT. Successful synthesis or integration of these efforts will be necessary to the overall success of the CTBT monitoring efforts. Hopefully, the reader will come away with an appreciation for the wide variety of problems and approaches to solutions currently underway within the DOE program.

◆ Challenges in Data Processing for CTBT Verification

The verification of a CTBT presents some significant challenges in the Automated Data Processing (ADP) arena. These challenges are driven by the lower event thresholds required by the CTBT, and they cover a wide variety of problems that range from increases in data volumes and types to the complications added by trying to integrate new sensor technologies and techniques into the framework of existing verification data processing systems. This section will briefly touch on some of the challenges in the automated data processing research area.

Central among the challenges is the increase in data volumes brought on by the lowered thresholds required by the CTBT. In general terms, the raw data volumes are expected to increase by an order of magnitude over current monitoring systems to roughly 10 Gbytes of data every day. This increase in raw data volume ripples through the entire data processing pipeline since it implies an increase in the number of stations to process, the number of detections generated at each station, the number of events formed by the system, etc. In addition to the demands placed on physical resources such as disk space, network bandwidth, and I/O channels, this increased data load also impacts software algorithms since performance requirements prohibit the use of algorithms that are not efficient with large quantities of data. CTBT-level data volumes also have implications for the work done by human analysts in the processing sequence. The number of events and the number of stations capable of being used in event formation will both be several times greater than they are today. Since budgets are unlikely to allow an increase in staff size, the automated systems for CTBT monitoring must become more accurate, or the analysts must become more efficient, or both.

In addition to the increase in data volumes from lowered thresholds, the International Monitoring System (IMS) network will include data from multiple sensor sources such as infrasound and radionuclide sampling sensors. These additional technologies will mean new algorithms and new problems unique to processing data from these sensor systems when compared to existing seismic monitoring systems. It is expected that the experience with seismic systems can be leveraged to help with these additional technologies, but unique challenges will continue to arise from the integration of sensor data from multiple technologies. In addition, new display and analysis tools may be required to take full advantage of this integrated data set.

Integration of sensor technologies also allows opportunities for synergy between sensor systems. In the past, sensor systems were essentially dedicated to a particular domain for monitoring purposes. Under a CTBT, information will be used from multiple sensor systems to fully understand certain events and to defeat cer-

tain evasion scenarios. How to integrate and exploit this synergy between systems remains a significant challenge for researchers in automated data processing and other areas.

Another chief challenge in the ADP arena comes from the need to incorporate regional knowledge about the Earth in order to accurately detect, locate, and identify events at CTBT thresholds. The research to develop this regional knowledge is a significant part of the overall DOE research program, but once acquired, serious challenges exist in terms of organizing, storing, and making this data available to automated processing routines. This task is complicated by the fact that this knowledge is available at differing resolutions over the Earth, and it is also recognized that the types and level of knowledge will change over time.

Finally, all of the research done to meet the above challenges must be done with the goal of integrating the solutions into the existing prototypes being developed for the US National Data Center (NDC) and the International Data Center (IDC). These prototypes are complex, evolving systems in their own right, and integration of new algorithms and techniques must not interfere with the development of the centers. Both the IDC and NDC are establishing testbeds and procedures to facilitate the integration process, but the need for integration of prototypes into the NDC and/or IDC environment complicates the development of research prototypes.

◆ ADP as an Integrating Technology

Automated data processing technology acts as the focal point for the synthesis or integration of the various sensing technologies used to monitor a CTBT. It provides a vehicle for examining the similarities between technologies and the tools needed to process the data from those technologies. It provides leverage for bringing new sensing technologies on-line quickly due to the ability to reuse algorithms across technologies. These attributes make the ADP arena an ideal place to explore synergies between technologies and the application of existing techniques to new technologies, or the application of new techniques to existing technologies.

In addition to its key role of synthesis across technologies, ADP also acts as a bridge or migration route between research and operations. In many cases, research remains unused or under-used because the results are often reports or other outputs that are not directly suitable, or at least logically extensible, to the operational environment. Because of the need to integrate with the existing processing environment, a portion of the effort in the ADP area must focus on the

space between research and operations. Work in the ADP area is truly applied research, and as such, is ideally suited to aiding the transition of other research results into the operational arenas.

◆ ADP Research within the DOE's CTBT R&D Program

Although research applying to ADP problems is ongoing at a number of government agencies, universities, and commercial companies, this paper will focus on the work within the DOE sponsored CTBT R&D program. The work in this program is divided into three main areas; advanced processing technology, computer-human interface technology, and information systems technology. For each area, a general overview of the work in that area will be presented along with some examples of efforts in this area. ADP is a broad ranging task area, however, and this paper does not pretend to cover the topic in-depth. The reader is encouraged to examine the other papers and presentations at the symposium for more information.

◆◆ Advanced Processing Technology

This task area focuses primarily on improvements to the automated engines that extract information from raw data. Within this task area, research is going into the development of new algorithms, the improvement of processing techniques using new computational technologies, and the exploration of cross-sensor synergies. As examples from within this area, the following paragraphs will briefly touch on research aimed at improving automated location capabilities, a method for doing full network event detection, and work being done to develop a high-level cross-sensor model of the overall CTBT network.

After careful consideration and consultation with the operational organizations, the decision has been made to place a priority on research into improved automated location techniques, especially those capable of improving depth estimation. Location is a strong indicator of event identity, and accurate locations are often a key to further processing necessary to refine an event. Several different directions are being investigated at the DOE labs to improve location capability. The first will focus on adaptive network locations that work to improve station corrections in regions with unknown velocity models. Another effort will address improving location capability by using a combination of travel time tables and waveform correlation techniques. Yet another effort will examine the problems associated with accurate location of individual events within a swarm. All of these efforts will result in new algorithms or techniques which can be applied to software in order to improve its capabilities.

Another area that has been the focus of considerable effort within the past few years is the issue of association of detections into events. A DOE-sponsored effort is underway to attempt direct event detection using the full data from a network of stations. This project (the Waveform Correlation Event Detection System - WCEDS) uses a uniform grid across the Earth's surface and into the subduction zones as search points. For each search point, the waveforms from all the stations are processed and aligned as if an event had happened at that point. The waveform pattern is then correlated with a master pattern for events at that location, and if the correlation exceeds the threshold, then an event is declared. This technique has the advantages of using all of the arrivals within the waveform to form the event, scaling well to larger numbers of stations, and being adaptable to distributed or parallel computer architectures. Early results in this effort have been promising, but this is clearly a longer term effort in order to produce a stable, reliable algorithm.

In a very different vein, work is also underway to develop a high-level model of the overall CTBT network. This model (the CTBT Integrated Verification System Evaluation Model - IVSEM) consists of integrated high-level models of seismic, hydroacoustic, infrasound, and radionuclide networks and can be used to evaluate the overall system performance of different numbers and types of sensors. It is intended as an affordable, portable model that is easy to use and understand, and it is envisioned as an aid for the treaty negotiation process. The model is designed to run on a portable 80486 or Pentium class machine, and it provides graphical outputs of its results as maps and charts. At this point, the model is capable of providing estimates of the network's ability to detect events, but future work will be aimed at adding the ability to estimate location and identification capability of the network.

◆ Computer-Human Interface Technology

While the Advanced Processing Technology efforts are focusing at improving the ability of the processing pipeline to automatically deal with the increasing number of events, the Computer-Human Interface efforts are aimed at making the analysts more productive as they deal with events. Both of the examples in this area are focused on exploring possible display methods that will improve the flow of information to an analyst, thereby allowing the analyst to make better decisions in a shorter period of time.

The first example is an effort at improving analyst efficiency by changing the approach used to evaluate events. Currently, event analysis starts with an analyst looking directly at the signal from a sensor, or at least a pre-processed version of that signal. The increasing number and types of sensors makes this an increasingly difficult method of event analysis. If, instead, the analyst were able to look

at a display that provided information about an event at the correct level of detail for the decisions being made about the event, then significant performance improvements might be realized. Work is underway to develop prototypes of such a level of detail display. This would act as a top end for the tools currently in use by analysts, so it would not replace them, but rather would allow the analyst to only examine those events which truly need human attention.

Another effort is aimed at providing very high dimensionality information to the analyst in an easily grasped format. Leveraging off of work done for the intelligence community, work is underway to use multi-dimensional clustering techniques to take a large number of relationships between elements of events and map them to a 2 or 3 dimensional space. By comparing the current event to a large population of other, well-known events, it is hoped that insights into the event's character can be discerned by the position of the event within the cluster of points representing other events. If this proves to be true, then the analysts will have a powerful tool that will allow them to assess in seconds a number of relationships between events that would today take many hours of the analysts time.

➡ Information Systems Technology

The third main area of the ADP portion of the CTBT R&D program is Information Systems Technology. This area focuses on the information handling and management infrastructure needed to allow the high-fidelity processing of the large volumes of data expected in a CTBT monitoring system. Examples in this area include the effort to develop a CTBT Knowledge Base to provide organized storage of the information needed by the ADP routines, and the efforts directed at data surety analysis.

A primary fallout of the move to lower thresholds in the CTBT environment is the need for detailed regional knowledge, such as travel time tables, to allow accurate locations for regional and local events. While this knowledge is being acquired in other portions of the CTBT R&D program, the task of developing a framework for the storing and retrieval of this knowledge is a task that falls within the ADP realm. The mechanism for providing this organized storage is the development of a CTBT Knowledge Base.

The Knowledge Base is envisioned as a storage area for the quasi-static parameters and geophysical data needed by the ADP routines. It will contain path-dependent information such as regional travel time tables, algorithmic information such as filter and beam sets, geophysical information for such as density and velocity models, and metadata to allow tracking of the knowledge both through time and the processing pipeline. One of the problems facing the monitoring sys-

tems today is the large number of ad-hoc mechanisms used to store knowledge today. This widely dispersed method of knowledge storage makes fine tuning of the system difficult, time consuming, and requires a great deal of familiarity with the whole system before a person can begin trying to fine tune. Another benefit of the Knowledge Base, therefore, will be its ability to consolidate the ad-hoc knowledge storage used by the current operations prototypes and improve the ease and accuracy of tuning the overall system.

The knowledge base is currently in the conceptual phase of development. A proposed Conceptual Requirements Document is available, and an effort is underway to fully identify the scope of the knowledge base and the types of data to be stored in it. That effort is expected to be complete soon, and the design process can then be undertaken.

Global monitoring systems clearly store a large quantity of information that would be a tempting target for tampering or destruction. Users place confidence in all types of data within the system from raw sensor data to knowledge base information and need confidence in its integrity and authenticity, so this data must be protected. At the same time, easy access to needed information is important for the participants. The efforts in the data surety area are balancing these requirements and making recommendations for future direction in this area.

◆ Summary

Monitoring a CTBT presents a number of significant challenges in the ADP area, and these challenges must be met with a variety of techniques and technologies. Success in the ADP area is crucial to the ability to monitor a CTBT, however, so successful synthesis of the various components within ADP should be a key goal for researchers everywhere.